Análise de séries temporais

por similaridade e alinhamento não linear com Dynamic Time Warping













Diego Furtado Silva



Rafael da Costa Silva



Luiz Antonio Rozendo

[2007-2011] Bacharelado em Ciência da Computação - ICMC-USP

[2011-2017] Mestre e Doutor em Ciência da Computação e Matemática Computacional - ICMC-USP

Com estágios de pesquisa na Columbia University e University of California, Riverside









[2007-Atual] Professor Adjunto no Departamento de Computação da Universidade Federal de São Carlos

diegofs@ufscar.br







Slide pros alunos falarem:D

Apresentação - curso

O que esperar do curso

- Análise de séries temporais (em largura)
- Um pouco de aprofundamento em técnicas baseadas em similaridade () Tw)
- Poucos exemplos em geofísica ou exploração e produção de óleo e gás

Vamos sempre falar um pouco, mostrar um pouco de código e passar exercícios

O que é uma série temporal?

Série temporal

Grosso modo:

- dados observados/coletados no tempo

Mas não pode ser só isso!

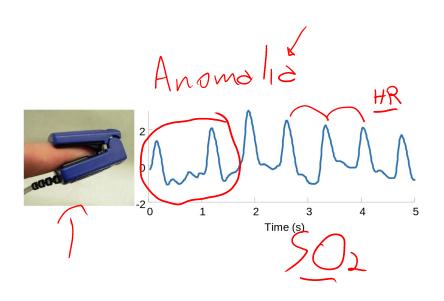


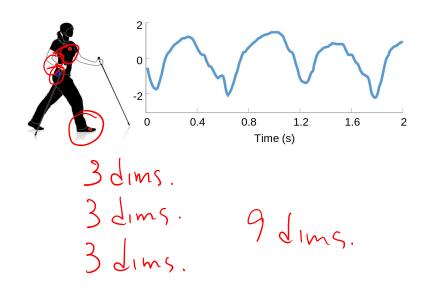
Uma série temporal é um **conjunto de valores ordenados** (x, tal que:

•
$$x = (x_1, x_2, ..., x_n)$$
• $x_i \in \mathbb{R}, \forall i \in \{1, 2, ..., n\}$

Ou seja, não precisa haver "variação no tempo"

p.s. Geralmente, considera-se espaçamento uniforme







Aizenberg, I., Sheremetov, L., Villa-Vargas, L., & Martinez-Muñoz, J. (2016). Multilayer neural network with multi-valued neurons in time series forecasting of oil production. Neurocomputing, 175, 980-989.

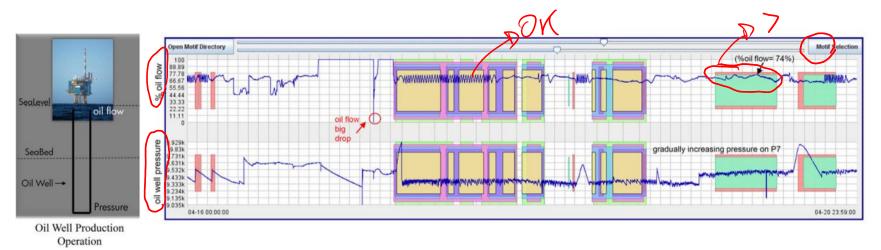
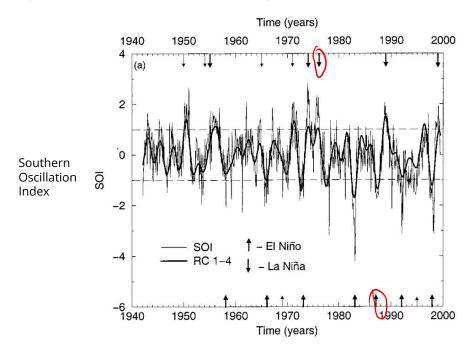
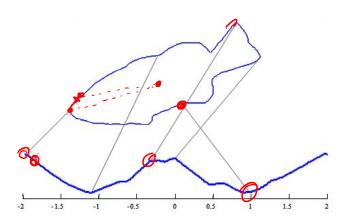


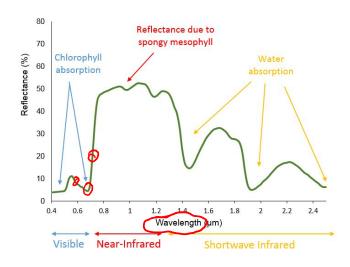
Figure 8. Oil well production time series with seven different frequent patterns with distinct colors (x-axis: time; y-axis: percentage oil flow and pressure; green, high oil well production; yellow, low oil well production).

Hao, M. C., Marwah, M., Janetzko, H., Dayal, U., Keim, D. A., Patnaik, D., ... & Sharma, R. K. (2012). Visual exploration of frequent patterns in multivariate time series. Information Visualization, 11(1), 71-83.



Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., ... & Yiou, P. (2002). Advanced spectral methods for climatic time series. Reviews of geophysics, 40(1), 3-1.





Essas eram séries temporais univariadas. Generalizando:

Uma série temporal é um **conjunto de valores ordenados x**, tal que:

•
$$x = (x_1, x_2, ..., x_n)$$

•
$$x_i \in \mathbb{R}^{n} \forall i \in \{1, 2, ..., n\}$$

$$: d = 1$$

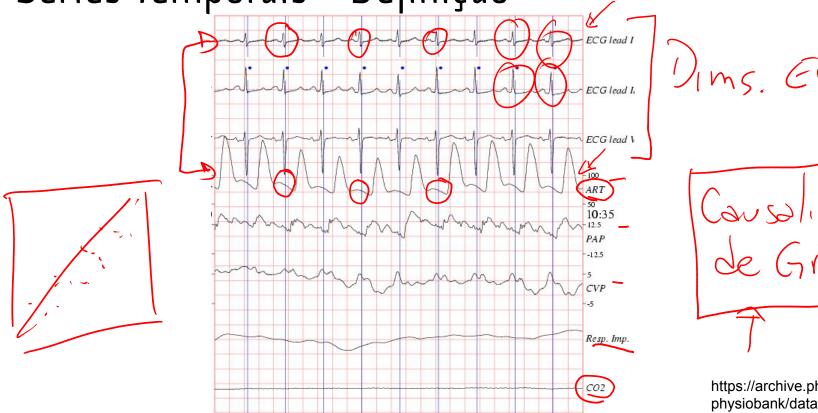
Univariada: d = 1

- Multidimensional vs Multivariada

Dependência

Acelevametro » dims. dep.

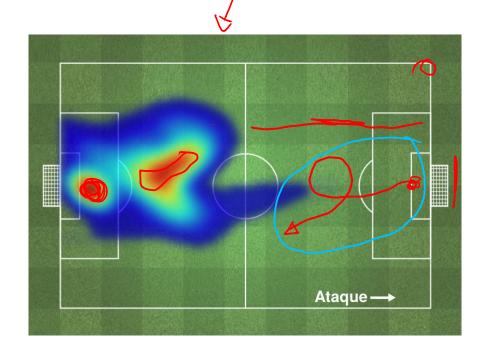
Ex. do paça » pressão e fluxo » indep.

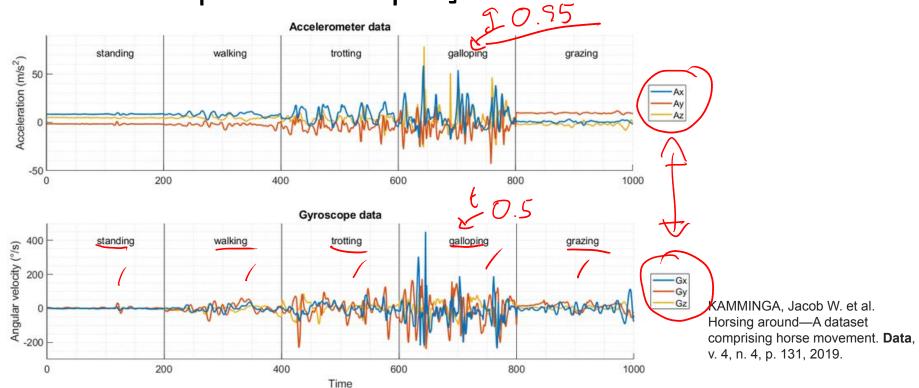


Dims. ECG

https://archive.physionet.org/ physiobank/database/mghdb/







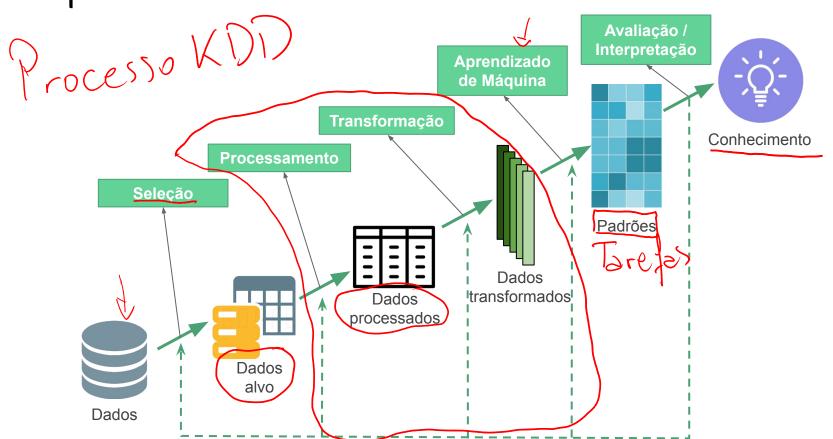
- Multidimensional vs Multivariada

- Usualmente, os algoritmos são propostos para uma dimensão
 - A adaptação para multi geralmente segue duas linhas
 - Dependendência vs independência das dimensões

e Séries Temporais

Mineração de Dados

Etapas da descoberta de conhecimento



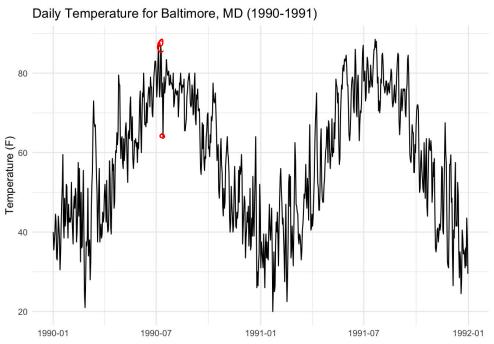
Do ponto de vista de "data analytics" analysis postprocessing preparation preprocessing visualization planning cleaning interpretation filtering data collection correlation documentation feature generation completion evaluation regression data selection correction forecasting standardization \ classification transformation 4 clustering

Fig. 1.1 Phases of data analysis projects

Pré-processamento (filtragem)

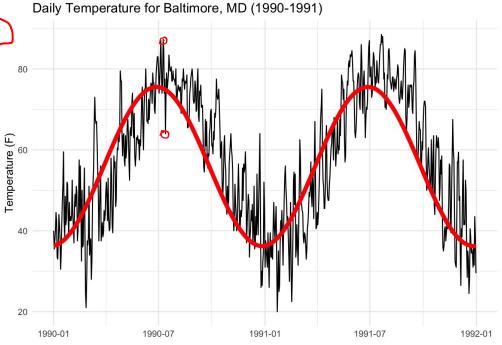
Ruídos (?)

ou complexidade



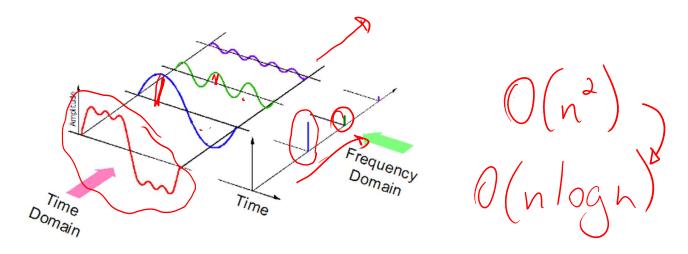
Ruídos (?)



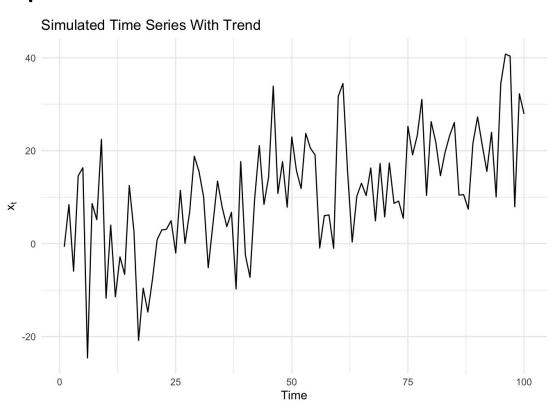


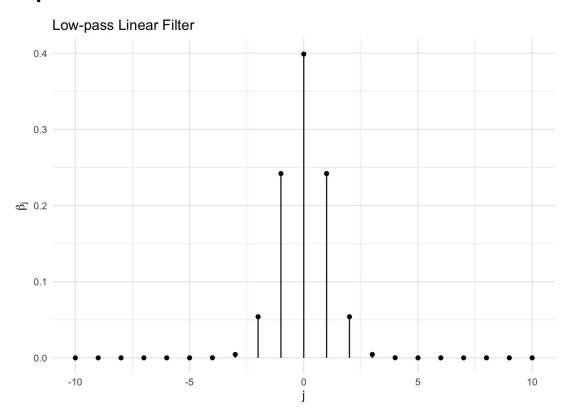
Fourier - tempo-frequência

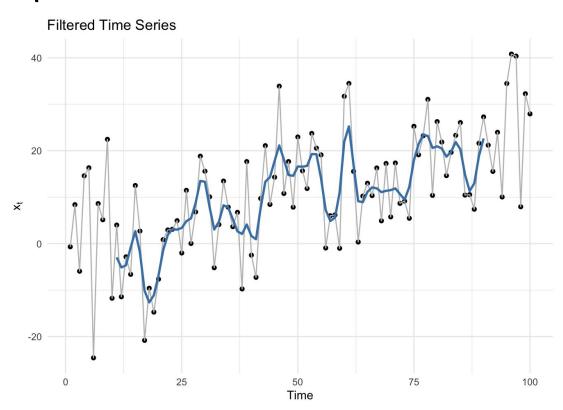
Sem se aprofundar

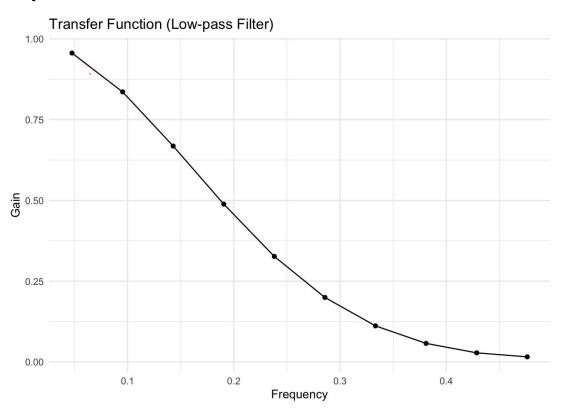


p.s. muitos consideram a FFT algoritmo mais importante de todos os tempos

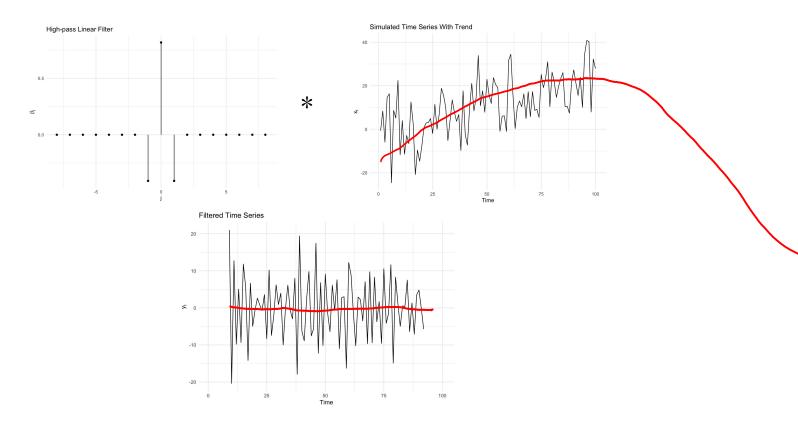








Filtro Passa-alta



Vamos à prática?

Mais filtros já com código e tudo, com o Rafael